

CLAIMS

What is claimed is:

1. A method for obtaining one or more candidate proteins for crystallization from a broad diversity sample, wherein the candidate proteins have desired characteristics to facilitate crystallization, the method comprising:
 - a) obtaining a broad diversity sample comprising microorganisms potentially having genes coding for one or more proteins having desired characteristics that facilitate crystallization;
 - b) isolating nucleic acids from the sample;
 - c) sequencing a plurality of nucleic acid segments comprised in the isolated nucleic acids;
 - d) selecting from the obtained nucleic acid sequences one or more target sequences based on suitable selection criteria;
 - e) optionally obtaining from the broad diversity sample one or more additional nucleic acid segments comprising the one or more target sequences or a part thereof, wherein the additional nucleic acid segment codes for the candidate protein or a part thereof;
 - f) expressing said one or more target sequences and/or additional nucleic acid segments; and
 - g) isolating expressed gene product(s) to obtain one or more candidate proteins that have characteristics that facilitate crystallization.
2. The method of claim 1, wherein the candidate proteins have desired characteristics to facilitate the process of structure determination.
3. The method of claim 1, wherein the suitable selection criteria comprise one or more criterion selected from the group consisting of:

- 5 a) a predetermined maximum hydrophobicity of any given region of a predetermined length of the sequence;
- b) a predetermined minimum percentage of one or more predetermined amino acid residues;
- 5 c) a predetermined maximum percentage of one or more amino acid residues;
- and combinations thereof.
4. The method of claim 3, wherein the suitable selection criteria comprise a criterion of a preselected minimum percentage of one or more amino acid
- 10 residues selected from the group consisting of Asn, Gln, Glu, Asp, His, Lys and combinations thereof.
5. The method of claim 3, wherein the suitable selection criteria comprise a criterion of a preselected maximum percentage of one or more amino acid
- 15 residues selected from the group consisting of Phe, Tyr, Trp and combinations thereof.
6. The method of claim 1, wherein the plurality of nucleic acid segments is selected such that it comprises nucleic acid segments suspected of coding for a protein or part of a protein of interest.
7. The method of claim 6, wherein oligonucleotide primers, derived from known
- 20 sequences coding for a proteins from the selected protein family of interest, are used in sequence-based screening methods using polymerase chain reaction (PCR) to select the plurality of nucleic acid segments
8. The method of claim 1, wherein the plurality of nucleic acid segments is comprised of a metagenomic gene library.

9. The method of claim 1, wherein the one or more candidate protein is a thermostable protein.
10. The method of claim 1, wherein the obtained sample comprises microorganisms selected from the group consisting of: viruses, prokaryotic microorganisms,
5 lower eukaryotic microorganisms and combinations thereof.
11. The method of claim 1, wherein the broad diversity sample is obtained from isolated strains of microorganisms.
12. The method of claim 11, wherein the microorganisms are thermophilic organisms.
- 10 13. The method of claim 1, wherein the broad diversity sample is obtained from a natural environment.
14. The method of claim 13, wherein the environment is a geothermal environment.
15. The method according to claim 13, wherein the broad diversity sample is enriched for a microbial population, prior to isolating nucleic acids, by
15 a) maintaining the sample under conditions substantially similar to the environment from which the sample was obtained to thereby expand the microbial population; and
b) allowing a sufficient quantity of a microbial population to expand; whereby the population has been enriched.
- 20 16. The method of claim 15, wherein the nucleic acids are biologically normalized by combining different enriched microbial populations prior to extracting the nucleic acids.

T0790" 22484800

17. The method of claim 7, wherein the primers are designed to preferentially screen and amplify candidate sequences from the protein family of interest that have one or more selected features.
18. The method of claim 2, wherein the suitable selection criteria benefit structure determination.
19. The method of claim 18, wherein the suitable selection criteria comprise a criterion of a desired number or ratio of a pre-determined amino acid residue.
20. The method of claim 19, wherein said criterion is a desired ratio of methionine residues.
21. The method of claim 6, wherein the candidate protein comprises an active site of a protein family.
22. The method of claim 6, wherein the protein family comprises a protein in a pathogenic organism.
23. The method of claim 6, wherein the protein family comprises a mammalian protein, including a human protein, with unknown structure.
24. The method of claim 23, wherein the mammalian protein with unknown structure is linked to a disease.
25. A method for obtaining a crystallized protein, comprising:
- a) obtaining a candidate protein using the method of claim 1; and
 - b) crystallizing said candidate protein.

26. A method for obtaining a three-dimensional structural information of a protein from a selected protein family, comprising
- a) obtaining a crystallized protein according to claim 25;
 - b) collecting diffraction data for the obtained crystal of the candidate protein;
 - c) optionally obtaining complementary data for phase determination of the diffraction data; and
 - d) determining the protein structure by use of the obtained data.
27. The method according to claim 26, wherein the protein structural information is used to facilitate protein design.
28. The method of claim 27, wherein the obtained plurality of nucleic acid sequences allows the determination of important functional determinants for designing proteins of new and/or improved functionality according to selected criteria.
29. The method of claim 28, where the new and/or improved functionality is achieved by rational design.
30. The method of claim 28, wherein the new and/or improved functionality is achieved by methods of directed evolution focusing on important amino acids or protein regions of importance for desired properties.
31. The method according to 26 claim wherein the structure information facilitates the design of a drug compound for combating a pathogenic organism.
32. The method according to claim 26, wherein the structure information facilitates the design of a therapeutic compound.

33. The method of claim 26, wherein selenomethionine is incorporated in the candidate protein.
34. The method according to claim 26, wherein the structural information becomes part of a database comprising structural information.
- 5 35. The method according to claim 26, wherein the structural information is used for structure prediction of proteins.
36. A method for obtaining the protein structure of a first protein from protein structure data which has insufficient phase information for a structure determination, comprising:
- 10 a) obtaining a protein structure of a second protein from the same protein family with the method of claim 26;
- b) determining the phase information for said structure data for said first protein with molecular replacement methods based on the obtained structure of said second protein; and
- 15 c) determining the protein structure by use of the initial structure data and the obtained phase information.
37. A method for predicting the structure of a first protein, comprising:
- a) obtaining a protein structure of a second protein from the same protein family with the method of claim 26; and
- 20 b) predicting the structure of said first protein with homology modeling based on the structure of said first protein.